Course Information

Course name: Stat 385: Methods for Data Analysis
Section: 002
Term: Spring 2019
Meeting times:

- Lecture: T, R 4-5:15pm in Duncan Hall 1042
- Lab: T 6-8pm KWG 130

Contact Information

Instructor: Julia Schedler
Office: Duncan Hall 1039 (on the first floor across from the mail room)
Email: julia.schedler@rice.edu
Office Hours:Mondays 6-8 or by appointment

TA: Violet Allen-Glass
Email: violet.m.allen-glass@rice.edu
Office Hours

- Time: TBD
- Location: TBD

Labbies:

- Catie Tsai (tsai@rice.edu)
- Niki Parekh (niki.parekh@rice.edu)

Course Description

From the rise of tech giants such as Google, Facebook, and Amazon to the success of teams like the Oakland Athletics, it is increasingly the case that businesses must keep up with the state of the art in statistical analysis to compete. Perhaps the most public evidence of the need for advanced analysis methods was in October 2006, when Netflix announced a competition (Links to an external site.)Links to an external site. to improve upon their content suggestion algorithm. Three years later, the winning team was awarded a 1 million dollar prize.
The darker side of this movement to leverage the massive amounts of data has shown up more recently. One issue is simply that large stores of personal data on millions of people are targets for hacking. This has happened so many times that there are websites such as Have I been Pwned? (Links to an external site.)Links to an external

site. which will tell you if your information (based on an email address) has been compromised in any hacks. A more insidious example of the risks of working with big data came into focus after the 2016 election, when it became clear that Russia-linked Facebook ads had targeted users in Wisconsin and and Michigan (Links to an external site.)Links to an external site.. This discovery has made the general public more aware of the inherent power of these platforms, and has led to calls of greater accountability for tech giants.
In order to run a successful (and ethical!) business, one must be aware of both the need for advanced data analysis as well as the inherent risks of said analysis. In this course, we'll cover the basics of more advanced kinds of statistical analyses as well as the ways they can be used to mislead. My goal for this course is to not only give you the technical knowledge you need to understand complex statistical methods, but also to give you the communication and critical thinking skills that will allow you to use that knowledge appropriately.

Learning Goals

- **Language** Statistics, just like any other discipline, has its own set of definitions and jargon. We will review the basics (mean, median, central tendency, variability, confidence…) and dive into the more advanced stuff (multivariate techniques, regression, dependent data) to build up your statistical vocabulary throughout the semester
- **Formulating Questions** We will learn the way to ask questions that can be addressed by the collection and analysis of data, as well as how to change a general question into a statistical one.
- **Visualization** We will discuss the components of a “good” visualization, learn how to create one, and common errors in visualization of data.
- **Theory** You should know from your prerequisite the basic theoretical properties of probability distributions, random variables, hypothesis tests, and confidence intervals. We will build on this knowledge to form theoretical understanding of more advanced (and more applicable) statistical techniques such as linear regression, ANOVA, multivariate techniques, etc.
- **Analysis** We will explore the pillars of a valid statistical analysis, including, asking the right question, visualization, choice of procedure, interpretation of results, and generalizability by framing it in the context of making good business decisions.
- **Communication** How do we talk about statistics? Which tools do you need to talk to a statistician? We will explore some commonly used advanced statistical procedures and practice discussing results of analysis methods we will be covering theoretically.
- **Skepticism** It's very easy to believe something when it comes from a trusted source. There is nothing wrong with this, but it is essential to know when and how to be skeptical of the results of a statistical analysis.

Required/Recommended Materials

The only book I am requiring is *Naked Statistics: Stripping the Dread from the Data* by Charles Wheelan. You can purchase this on Amazon for around $12 (Links to an external site.)Links to an external site..

You can also purchase the textbook I will be loosely following, but it is by no means required:

Data Analysis & Decision Making (Links to an external site.)Links to an external site.

Author: Albright

Edition: 4th

ISBN: 9780538476126

Copyright Year: 2011

Publisher: Cengage Learning

Course Content

Date	Content	Lab	Reading to be completed for in-class discussion
Tuesday, January 8	Introductions/Review	Installing VM software	
Thursday, January 10	Review		Introduction
Tuesday, January 15	Review	Review Topics	Chapter 1
Thursday, January 17	Linear Regression		
Tuesday, January 22	An Ethical Crisis in Computing?	Linear Regression	Chapter 2
Thursday, January 24	Linear Regression		
Tuesday, January 29	Linear Regression	Homework help	Chapter 3

Thursday, January 31	ANOVA		
Tuesday, February 5	Design of Experiments	ANOVA/DOE	Chapter 4
Thursday, February 7	Midterm recess- no class		
Tuesday, February 12	Generalized Linear Models- logistic regression	GLM	
Thursday, February 14	Generalized Linear Models- Poisson regression		
Tuesday, February 19	Process Control/Optimization	Process Control/Optimization	Chapter 5
Thursday, February 21	Process Control/Optimization		
Tuesday, February 26	Nonparametric Approaches	Simulation	Chapter 6
Thursday, February 28	Nonparametric Approaches		
Tuesday, March 5	Bayesian Approaches	Bayesian Approaches	Chapter 7
Thursday, March 7	Bayesian Approaches		
Tuesday, March 12	SPRING BREAK!		
Thursday, March 14	SPRING BREAK!		
Tuesday, March 19	Multivariate Statsitics	Multivatiate	Chapter 8,9

Thursday, March 21	Multivariate Statistics		
Tuesday, March 26	Multivariate Statistics	Multivariate, ctd	Chapter 10
Thursday, March 28	Dependent Data		
Tuesday, April 2	Dependent Data	Dependent Data	Chapter 11
Thursday, April 4	Wiggle Room		
Tuesday, April 9	Wiggle Room	Presentation prep	Chapter 12
Thursday, April 11	Presentations		
Tuesday, April 16	Presentations		Chapter 13
Thursday, April 18	Presentations		Conclusion

Course Structure

"Lectures" will consist of fill in the blank notes to facilitate time management. Research says that 18-22 year olds have about a 10 minute attention span, so I will try not to talk for more than 10 minutes continuously. The format will be 10 minutes or so of me introducing a new concept/doing an example, then you will work on examples in groups, then we will review and repeat the process as many times as we can during the allotted time.

We will also have in-class discussions over reading in the required book *Naked Statistics: Stripping the Dread from the Data*. There will be “reading quizzes” you will be required to complete prior to our in-class discussions. Participation in discussions will count towards your participation grade!

Lab Session

Labs are an opportunity for you to work on homework. Labbies will take attendance and this will count towards your participation grade.

Assignments

- Homework
- Reading Quizzes

- Canvas Discussions
- Project: This will be an individual data analysis project which will involve a typed report as well as a 10 minute in-class presentation. Details will be provided towards the middle of the semester.

Exams

We will have a midterm exam and a final. The midterm will be take-home and will be distributed Thursday, February 21 and collected Tuesday, February 26 and will cover material up to and including whatever we have covered as of Thursday, February 14 (i.e. it will not include material covered on the 19th and 21st).

The final exam will be held whenever/wherever the Office of the Registrar indicates. If you have conflicts, we can discuss them as soon as we know when the final is.

Grading

- **Homework (30%)** - homework will be graded for both completeness and correctness. I will drop your lowest score.
- **Midterm Exam (15%) -** the midterm will be graded for both completeness and correctness. Show as much work as possible to receive maximal partial credit. See the exam section for details on the date and content coverage.
- **Final Exam (15%) -** the midterm will be graded for both completeness and correctness. Show as much work as possible to receive maximal partial credit. If your final exam score is higher than your midterm score, I will replace your midterm score with your final exam score.
- **Projects (25%) -** the project will be graded based equally on both the written report and the in-class presentation. I will provide detailed rubrics for both parts.
- **Reading Quizzes (5%) -** the multiple choice questions will be graded based on correctness. The short answer questions will be graded for completion, but you need to make an honest attempt at an answer.
- **Participation (10%) -** your participation grade will come from lab attendance (2%), class attendance (3%), and your participation in in-class discussions (5%). For the attendance portions you only need to show up; to receive full credit for the in-class discussions you must speak in at least 60% of the discussions (I'll be keeping track).

Late/Make-up work policy

- Quizzes: barring a university excused absense/family emergency, I will not accept late quiz submissions. I will drop your two lowest scores, so you can miss two quizzes with no consequence.
- Homework: Late homework will be accepted, with 10% deducted from your final score for each day it is late (e.g. if you turn in homework 1 by 11:59 PM on Jan 21, you will be able to score a maximum of 90%). If you do turn your homework in late, please do not look at the solutions before you have submitted. If you have turned the

homework in on time, please do not share the solutions with anyone who has not yet turned it in.
- Exams: The midterm is take-home, so there should not be any reason you need to make it up since you will have a week to complete it. If you have a university excused absence during the exam time, we can discuss extending the deadline for turning it in. If you have a conflict with the final exam time, let me know as soon as possible so we can work something out.

Rice Honor Code

In this course, all students will be held to the standards of the Rice Honor Code, a code that you pledged to honor when you matriculated at this institution. If you are unfamiliar with the details of this code and how it is administered, you should consult the Honor System Handbook at http://honor.rice.edu/honor-system-handbook/. This handbook outlines the University's expectations for the integrity of your academic work, the procedures for resolving alleged violations of those expectations, and the rights and responsibilities of students and faculty members throughout the process.

Disability Resource Center

If you have a documented disability or other condition that may affect academic performance you should: 1) make sure this documentation is on file with the Disability Resource Center (Allen Center, Room 111 / adarice@rice.edu / x5841) to determine the accommodations you need; and 2) talk with me to discuss your accommodation needs.

Consent for study to be completed on data from this class

This semester your instructor will be conducting a study looking at teaching and learning. The purpose of this study is to determine the factors that influence teaching effectiveness and learning. Your participation in this study will last for the duration of the current semester and will entail activities no different from the regular activities you would otherwise engage in as part of the course. If you would like more information, please do not hesitate to ask. If you would like to opt out of the study, please talk to me as soon as possible.

Specifically, you will be answering some survey questions both at the beginning and end of the semester to measure your skepticism. We will analyze the data as a class at the end. There will also be a few reflection questions to be answered at the end of the semester. I will be writing a paper discussing teaching skepticism in a statistics course. All data included in the paper will be anonymous, and selected comments from reflection questions will only be included if you wish them to be.

Syllabus Change Policy

This syllabus is only a guide for the course and is subject to change with advanced notice.